

Les dades de recerca: una oportunitat professional per als gestors d'informació

Llorenç Arguimbau

Observatori de la Recerca (OR-IEC). Institut d'Estudis Catalans (IEC)

llarguimbau@iec.cat

Resum:

En un context d'obertura de quantitats massives de dades, aquestes són considerades com actius amb valor afegit. Per tant, les dades científiques s'han de gestionar, compartir i preservar a llarg termini, des de la seva creació fins a la conservació definitiva, i mantenir l'enllaç amb les publicacions científiques derivades. L'article s'estructura en cinc apartats: marc conceptual, infraestructures de dades, agents involucrats en la comunicació científica, documentació científica i gestió de dades, i experiències pràctiques (directoris internacionals i iniciatives espanyoles). Per acabar, s'extreuen algunes conclusions, tot i que es tracta d'una àrea emergent amb nombroses qüestions a resoldre, on els gestors d'informació hi poden jugar un paper destacat.

Paraules clau:

accés obert, ciència, comunicació científica, dades, dipòsits digitals, gestió, infraestructures, preservació, publicacions, recerca, reutilització

Los datos de investigación: una oportunidad profesional para los gestores de información

Resumen:

En un contexto de apertura de cantidades masivas de datos, éstos son considerados como activos con valor añadido. Por lo tanto, los datos científicos han de ser gestionados, compartidos y preservados a largo plazo, desde su creación hasta la conservación definitiva, y mantener el enlace con las publicaciones científicas derivadas. El artículo se estructura en cinco apartados, que corresponden a sus objetivos básicos: marco conceptual, infraestructuras de datos, agentes involucrados en la comunicación científica, documentación científica y gestión de datos, y experiencias prácticas (directorios internacionales e iniciativas españolas). Por último, se extraen algunas conclusiones, aunque se trata de un área emergente con numerosas cuestiones por resolver, donde los gestores de información pueden jugar un papel destacado.

Palabras clave:

acceso abierto, ciencia, comunicación científica, datos, repositorios digitales, gestión, infraestructuras, preservación, publicaciones, investigación, reutilización

Research data: professional opportunity for information managers

Abstract:

In the context of huge amounts of data becoming publicly available, this data is considered as an asset with added value. There should therefore be long term management dissemination and preservation of scientific data from its creation to its ultimate storage. Moreover, it should be referenced in scientific publications in which it is used. The article is divided into five sections which correspond to its basic objectives: conceptual framework; data infrastructures; agents involved in scientific communication; scientific documentation and data management; case studies (international directories and Spanish initiatives). Finally, some conclusions are drawn, all be it that this is an emerging field with various questions still to be addressed. It is in this context that information managers could play an important role.

Keywords:

open access, science, scientific communication, data, digital repositories, management, infrastructures, preservation, publications, re-search, reuse



©istockphoto/AlexRaths

Introducció

Les tecnologies de la informació i de la comunicació (TIC) i el moviment d'accés obert (*open access*) al coneixement han tingut un impacte evident en el món de la comunicació científica, sobretot en les publicacions. Però, què succeeix amb les dades creades o recollides per qualsevol investigació científica? Quin paper hi poden jugar els professionals de la gestió d'informació? Aquestes dues preguntes configuren l'eix central al voltant del qual gira l'article. Hem de tenir en compte que el tractament de dades científiques és una àrea emergent que presenta encara nombroses qüestions per resoldre.

«Durant els últims anys s'han incrementat extraordinàriament el volum i la rapidesa amb què aquestes dades [científiques] s'obtenen i es comparteixen gràcies al perfeccionament dels instruments de mesura i a la utilització

d'Internet per a la comunicació entre els investigadors. Com a resultat, la cadena de la comunicació científica s'ha ampliat [...] per incorporar la captura, descripció i reutilització de les dades primàries en què es basen aquestes publicacions»¹. Per tant, també cal administrar el cicle global de les dades, des de la creació fins a la conservació definitiva, i mantenir l'enllaç amb les publicacions científiques que se n'han derivat.

Hi ha un debat intens dins de la comunitat científica sobre la necessitat i els procediments per administrar, compartir i reutilitzar les dades². En els darrers anys, aquesta temàtica ha despertat l'interès de tots els agents involucrats en la comunicació científica: investigadors, institucions científiques, organismes internacionals, governs i agències financeres, editorials i bibliotecaris-documentalistes. Un article científic pot quedar obsolet, però no les dades en sí, perquè poden ser reutilitzades, recombinades o repro-

1. Àngel Borrego, «Enriquir les publicacions amb dades empíriques» [En línia], *Blok de BiD*, 11/04/2012. <<http://www.ub.edu/blokdebid/ca/content/enriquir-les-publicacions-amb-dades-emp%C3%ADriques-0>> [Consulta: 06/06/2013].
2. Daniel Torres-Salinas; Nicolás Robinson-García; Álvaro Cabezas-Clavijo, «Compartir los datos de investigación en ciencia: introducción al *data sharing*», *El Profesional de la Información*, vol. 21, núm. 2, p. 174.

cessades amb una altra tecnologia o simplement vistes des d'una altra perspectiva³.

Tot i que el fet de compartir dades entre els científics és consubstancial a la recerca, cal assenyalar algunes fites destacades de les darreres dècades. L'any 1971 es va crear Protein Data Bank (PDB)⁴ (Figura 1), el primer banc de dades científiques. Així mateix, l'any 1983, *Journal of Biological Chemistry* va ser una revista pionera a exigir les dades per a la publicació d'articles⁵. A partir de mitjans dels anys noranta, algunes agències anglosaxones encarregades del finançament de la recerca van començar a considerar que les dades s'han de compartir, sobretot si s'han obtingut amb diners públics. L'any 2003, la declaració de Berlín sobre l'accés obert al coneixement en ciències i humanitats contemplava també les dades.



Figura 1. Protein Data Bank (PDB)

L'article vol ser una introducció general al tema i, per tant, la metodologia està fonamentada en la revisió bibliogràfica i en la identificació de fonts d'informació de referència.

1. Marc conceptual

Quantitats ingents de dades són presents arreu de la nostra societat, mitjançant les aplicacions web 2.0, les administracions públiques, els mitjans de comunicació, les institucions científiques, etc. *Big data* és un concepte habitual que s'utilitza per definir grans paquets de dades, els més coneguts dels quals estan vinculats amb l'anàlisi de l'ús de xarxes socials i de motors de cerca a Internet. A la vegada, totes aquestes dades es troben cada cop més lligades als moviments *open*⁶. D'aquesta manera, creix la consciència que cal posar-les a disposició de tothom i preservar-les com a actius d'utilitat social i comuna⁷.

Segons l'Open Knowledge Foundation (OKFN)⁸, una dada oberta és qualsevol «dada que pot ser utilitzada, reutilitzada i redistribuïda lliurement per qualsevol persona, tan sols subjecte, com a màxim, a l'exigència d'atribució i de compartir en les mateixes condicions». Així, les propietats principals de les dades obertes són les següents:

- Disponibilitat i accés: han d'estar disponibles en el seu conjunt, amb un cost raonable de reproducció (preferentment, mitjançant la descàrrega a Internet) i en una forma adequada i modificable.
- Reutilització i redistribució: s'han de proporcionar en condicions que facilitin el reaprofitament, fins i tot la combinació amb altres conjunts de dades (*datasets*).
- Participació universal: tothom ha de ser capaç d'utilitzar, reutilitzar i redistribuir les dades obertes sense discriminar cap camp d'acció, persona o grup (per exemple, sense restriccions a usos comercials o limitacions per a finalitats educatives).

3. Remedios Melero, «Una plenamar de dades» [En línia], *Blok de BiD*, 24/11/2010. <<http://www.ub.edu/blokdebid/ca/content/una-plenamar-de-dades>> [Consulta: 06/06/2013].
4. Protein Data Bank (PDB) (www.rcsb.org/pdb) [Consulta: 06/06/2013].
5. Daniel Torres-Salinas, «Compartir datos (*data sharing*) en ciencia: contexto de una oportunidad», *Anuario ThinkEPI*, vol. 4 (2010), p. 259.
6. Antonia Ferrer-Sapena; Enrique A. Sánchez-Pérez, «Open data, big data: ¿hacia dónde nos dirigimos?» [En línia], *Notas ThinkEPI*, 1 febrero 2013. <<http://www.thinkepi.net/open-data-big-data-hacia-donde-nos-dirigimos>> [Consulta: 06/06/2013].
7. Alicia García-García; Josep-Manuel Rodríguez-Gairín; Tomás Saorín; Luis-Millán González; Xavi García-Massó; Antonia Ferrer-Sapena; Fernanda Peset, «ODiSEA: International Registry on Research Data» [En línia], *BiD: textos universitaris de biblioteconomia i documentació*, núm. 29 (2012). <<http://www.ub.edu/bid/29/garcia2.htm>> [Consulta: 06/06/2013].
8. OKFN, *The Open Data Handbook*, Cambridge: OKFN, 2012. <<http://opendatahandbook.org>> [Consulta: 06/06/2013].

Tim Berners-Lee⁹ ha proposat un esquema de cinc estrelles per a les dades obertes: s'atribueix una estrella a les dades en qualsevol format amb una llicència, i cinc estrelles a les dades enllaçables o *linked data*.

Pel que fa a la definició de dada de recerca, els National Institutes of Health (NIH) (Figura 2) dels Estats Units d'Àmerica afirmen que es tracta de qualsevol «material factual enregistrat que és acceptat de forma generalitzada per la comunitat científica per validar els resultats d'investigació»¹⁰. A la vegada, aquesta definició és el punt de partida a l'elaborada posteriorment per l'Organització per a la Cooperació i el Desenvolupament Econòmic (OCDE)¹¹. Ambdues organitzacions no consideren dades de recerca les notes de laboratori, les anàlisis preliminars, els esborranys d'articles científics o els projectes d'investigacions futures.

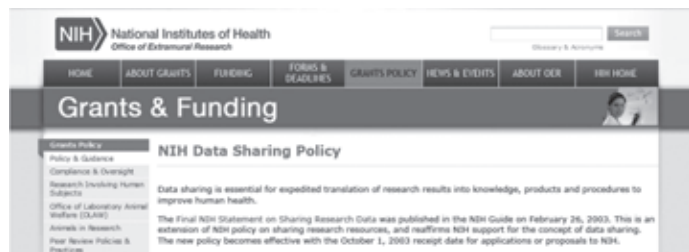


Figura 2. NIH Data Sharing Policy

Un informe de la Research Information Network (RIN)¹² ha elaborat una classificació de les dades de recerca a partir de tres criteris no excloents:

- Procés d'obtenció: experiments científics (per exemple, seqüències genètiques o experiments de física de partícules); simulacions (moviments astronòmics o models econòmics); o observacions de fenòmens específics (per exemple, enquestes sociològiques o dades climatològiques).
- Finalitat de recollida: projecte específic, d'abast mitjà o d'interès general.
- Tractament: dades primàries creades o recollides en experiments o observacions (*raw data*), o bé derivades a partir del tractament d'aquestes dades (*final research data*).

Ara bé, aquest marc conceptual no s'ajusta a totes les especialitats del coneixement: per exemple, un investigador d'humanitats pot treballar amb qualsevol registre fruit de l'experiència humana¹³. Per tant, el concepte de dada de recerca varia de manera considerable segons la disciplina científica i el mètode d'investigació. «La naturaleza de los datos de investigación es muchísimo más variable y dependiente de la disciplina y de su particular ciclo de vida. Además, los requisitos técnicos y legales para garantizar el acceso son más complejos que los de las publicaciones»¹⁴.

El *data sharing* consisteix a compartir en accés obert les dades científiques amb la finalitat que puguin ser reutilitzades per altres investigadors per generar coneixement. El potencial de la reutilització com a eina de validació de la ciència i com a motor d'innovació és inqüestionable¹⁵.

9. Tim Berners-Lee, *5 Star Open Data* [En línia]. <<http://5stardata.info>> [Consulta: 06/06/2013].

10. NIH, *NIH Data Sharing Policy* [en línia]. Bethesda, Maryland: NIH, 2007. <http://grants.nih.gov/grants/policy/data_sharing> [Consulta: 06/06/2013].

11. OCDE, *OECD Principles and Guidelines for Access to Research Data from Public Funding* [En línia], Paris: OCDE, 2007. <<http://www.oecd.org/science/sci-tech/38500813.pdf>> [Consulta: 06/06/2013].

12. Alma Swan; Sheridan Brown, *To Share or not to Share: publication and Quality Assurance of Research Data Outputs* [En línia], London: Research Information Network, 2008. <<http://www.rin.ac.uk/our-work/data-management-and-curation/share-or-not-share-research-data-outputs>> [Consulta: 06/06/2013].

13. Torres-Salinas; Robinson-García; Cabezas-Clavijo, *op. cit.*, p. 176.

14. Cristina González Copeiro; Jordi Serrano-Muñoz (coord.), *La conservación y reutilización de los datos científicos en España: informe del grupo de trabajo de buenas prácticas* [En línia], Madrid: FECYT, 2012. <http://www.recolecta.net/buscador/documentos/informe_datos_cientificos_en_esp.pdf> [Consulta: 06/06/2013].

15. Fernanda Peset; Antonia Ferrer-Sapena; Imma Subirats-Coll, «Open data y linked open data: su impacto en el área de bibliotecas y documentación», *El Profesional de la Información*, vol. 20, núm. 2 (2011), p. 165-173.

En aquest sentit, sembla evident que les dades obtingudes amb finançament públic haurien de ser compartides amb tothom. Nombroses agències de finançament i institucions científiques ja disposen de polítiques per impulsar l'accés obert al coneixement.

Amb el propòsit d'incentivar el moviment per compartir i reutilitzar dades de recerca, cal promocionar els seus beneficis:

- Progrés científic: nous estudis que combinen diversos conjunts de dades, els quals s'analitzen des de perspectives innovadores¹⁶.
- Eficiència: millor rendiment dels recursos econòmics invertits, eliminació de projectes duplicats i estalvi de costos¹⁷.
- Transparència: lluita contra el frau científic¹⁸.
- Visibilitat: augment del nombre de citacions dels treballs amb dades obertes¹⁹.

Malgrat els avantatges, el *data sharing* no és una pràctica generalitzada entre els científics. Segurament hi ha massa obstacles i incògnites que encara cal superar:

- Marc inestable: manca de reconeixement i protecció dels investigadors que comparteixen²⁰.
- Mal ús: utilització fraudulenta de les dades o crítica, més o menys justificada, a la investigació realitzada²¹.
- Capacitació insuficient: desconeixement i manca de

formació dels investigadors per garantir un tractament adequat de les dades obertes²².

- Dificultats tècniques: la implantació i gestió de polítiques i d'infraestructures s'enfronta a costos elevats, manca d'estandardització de les dades, formats múltiples, etc. En aquest sentit, els *datasets* poden ser objectes compostos i heterogenis, amb diferents elements o tipus de dades: textos, fulls de càlcul, gràfics, imatges, etc.²³

2. Infraestructures de dades

Les dades s'han d'administrar i preservar mitjançant infraestructures fiables i duradores. «De los numerosos procesos que tienen lugar durante la gestión de todo el ciclo de vida de los datos científicos, el almacenamiento en depósitos cobra una gran importancia por sus implicaciones en cuanto a la preservación para el futuro, la citación, el reconocimiento de la autoría, la utilización de enlaces persistentes...»²⁴.

La pregunta sobre com es poden gestionar, conservar i compartir dades científiques presenta diverses respostes. Una solució són els bancs de dades (*data banks*), infraestructures específicament dissenyades per a aquesta tasca i, per tant, l'objectiu ideal de qualsevol disciplina científica²⁵. Ara bé, també cal tenir en compte els dipòsits digitals²⁶, on les dades conviuen amb les publicacions i altres

16. A. Ramasamy; A. Mondry; C.C. Holmes; D.G. Altman, «Key Issues in Conducting a Meta-Analysis of Gene Expression Microarray Datasets» [En línia], *PLoS Medicine*, vol. 5, núm. 9, e184. <<http://dx.doi.org/10.1371/journal.pmed.0050184>> [Consulta: 06/06/2013].

17. García-García *et al.*, 2012, *op. cit.*

18. Torres-Salinas, *op. cit.*, p. 259.

19. Heather A. Piwowar; Roger S. Day; Douglas B. Fridsman, «Sharing Detailed Research Data Is Associated with Increased Citation Rate» [En línia], *PLoS ONE*, vol. 2, núm. 3, e308. <<http://dx.doi.org/10.1371/journal.pone.0000308>> [Consulta: 06/06/2013].

20. Christine L. Borgman, «Data, disciplines, and scholarly publishing», *Learned Publishing*, vol. 21, núm. 1 (2008), p. 29-38.

21. Torres-Salinas; Robinson-García; Cabezas-Clavijo, *op. cit.*, p. 181.

22. Torres-Salinas, *op. cit.*, p. 259.

23. González Copeiro; Serrano-Muñoz, *op. cit.*, p. 15.

24. González Copeiro; Serrano-Muñoz, *op. cit.*, p. 37.

25. Torres-Salinas; Robinson-García; Cabezas-Clavijo, *op. cit.*, p. 176.

26. Silvia Arano; Gemma Martínez; Marina Losada; Marta Villegas; Anna Casaldàliga; Núria Bel, «La comunidad "Recursos y datos primarios" de la Universitat Pompeu Fabra: los repositorios institucionales como infraestructuras científicas: estudio de caso», *Revista Española de Documentación Científica*, vol. 34, núm. 3 (2011), p. 387.

Per ara, els bancs de dades s'adrecen a comunitats científiques molt concretes: astronomia, ciències de la terra, gnòmica i proteòmica, neurociència, oncologia, etc.

materials, i les revistes²⁷, en les quals les dades actuen com a material suplementari dels articles. D'altra banda, els canals informals (peticions concretes als autors, llocs web d'investigadors o de grups de recerca, etc.) són una alternativa poc aconsellable perquè dificulten la localització i reutilització de dades degut a diferents factors: multiplicitat de formats, manca de normalització, desconeixement del context de creació, etc.

Similars als dipòsits de literatura científica, els *data banks* estan orientats específicament a la gestió de dades especialitzades en una disciplina científica. Per tant, es disposa d'una gran varietat de solucions i resulta difícil consensuar directrius o orientacions. En aquest sentit, són eines complexes per al dipòsit i recuperació de dades que requereixen d'especialistes i que provoquen costos elevats de manteniment (cal garantir-ne la viabilitat econòmica). Per ara, els bancs de dades s'adrecen a comunitats científiques molt concretes: astronomia, ciències de la terra, gnò-



©istockphoto/Denysov Dmytro

mica i proteòmica, neurociència, oncologia, etc. Per exemple, GenBank²⁸ és un model internacional en la publicació en accés obert de dades genètiques des de l'any 1982.

En relació als dipòsits digitals, generalment s'han orientat cap a la producció científica, deixant de banda formats menys habituals com els materials multimèdia, els programes d'ordinador o les dades²⁹. Com a prova d'això, el maig de 2013 el directori OpenDoar té registrats 2.279 dipòsits, dels quals tan sols 84 (un 3,7%) recopilen conjunts de dades.

Una altra possibilitat per compartir dades de recerca és fer-ho com a material suplementari als articles de revista. En aquesta línia, Borrego³⁰ comenta un informe del projecte europeu *Opportunities for Data Exchange* (ODE) que examina el paper de les editorials. L'informe identifica diversos mecanismes perquè les editorials facin disponibles els fitxers de dades vinculades amb els articles que publiquen, ja sigui mitjançant la publicació de materials

27. Borrego, 11/04/2012, *op. cit.*

28. GenBank (www.ncbi.nlm.nih.gov/genbank) [Consulta: 06/06/2013].

29. Arano *et al.*, *op. cit.*, p. 387.

30. Borrego, 11/04/2012, *op. cit.*

suplementaris (en el servidor de l'editorial o en un dipòsit extern) o, més recentment, mitjançant l'aparició de revistes dedicades a la publicació d'articles descriptius sobre conjunts de dades disponibles en un dipòsit (vegeu, per exemple, les revistes *Earth System Science Data* o *Giga Science*). Tanmateix, l'informe detecta dificultats rellevants perquè les editorials es responsabilitzin adequadament de la revisió, publicació i preservació de dades.

Pel que fa a les qüestions de caire més tècnic, el format i la descripció dels fitxers de dades són factors essencials per garantir la consulta i preservació a llarg termini³¹. És recomanable usar formats oberts i que permetin la indexació del contingut. També resulta necessari un control de versions dels fitxers, que han d'estar ben organitzats amb una estructura ordenada i una nomenclatura que permeti la identificació dels continguts. A més, la descripció mitjançant metadades facilita la localització, comprensió i utilització de les dades. A la vegada, les metadades s'estructuren en registres segons esquemes normalitzats, que són prioritaris per a la interoperabilitat dels sistemes. També cal tenir en compte que els conjunts de dades s'han d'associar a un identificador digital únic i permanent. Finalment, és recomanable fer còpies de seguretat en diferents tipus de suports físics, garantir la integritat dels fitxers i prevenir l'obsolescència tecnològica.

3. Agents de la comunicació científica

Una gran diversitat d'actors estan involucrats en el cicle de les dades científiques, des dels reguladors (governos, agències finançadores i organismes internacionals), pas-

sant pels gestors (editorials i biblioteques), fins als productors (investigadors i institucions científiques).

Des de mitjans dels anys noranta, cada cop més organismes encarregats del finançament de la recerca científica van començar a prendre consciència de la necessitat de compartir les dades, sobretot les que s'havien obtingut a partir de fons públics³². En aquesta línia, es dissenyaren polítiques, estàndards, directrius i recomanacions des d'entitats com, per exemple, l'Economic and Social Research Council (ESRC) del Regne Unit, la National Science Foundation (NSF) dels Estats Units o l'Australian National Data Service (ANDS). Degut al seu caràcter pioner, cal destacar la política dels NIH nord-americans, que des de l'any 2003 obligava a compartir les dades dels projectes d'investigació finançats amb diners públics. Així, les sol·licituds presentades a les seves convocatòries havien d'incorporar un pla de gestió de dades i, a més, els NIH impulsaren diversos dipòsits digitals. Per tant, les agències de finançament han de tenir un paper clau per garantir la sostenibilitat de les dades, a partir de l'exigència d'un pla de tractament i preservació³³.

Pel que fa als organismes internacionals, la UNESCO manifestà la necessitat de prendre mesures per a «encoratjar les universitats i altres organismes de recerca, tant públics com privats, per garantir la conservació de les dades d'investigació»³⁴. D'altra banda, l'OCDE impulsà el *data sharing*, mitjançant una guia per estimular l'accés a les dades procedents de finançament públic³⁵.

La Unió Europea (UE) considerarà les dades científiques com un factor de progrés econòmic, ja que la reutilització de dades prèvies proporcionava una oportunitat d'innovació a un cost raonable³⁶. D'aquesta manera, la directiva

31. González Copeiro; Serrano-Muñoz, *op. cit.*, p. 17-23.

32. Torres-Salinas; Robinson-García; Cabezas-Clavijo, *op. cit.*, p. 178.

33. Alice Keefer, «La preservació de les dades de recerca i les agències de finançament de la R+D» [En línia], *Blok de BiD*, 11/10/2011. <<http://www.ub.edu/blokdebid/ca/node/129>> [Consulta: 06/06/2013].

34. UNESCO, *Charter on the Preservation of Digital Heritage* [En línia], Paris: UNESCO, 2003. <http://portal.unesco.org/en/ev.php-URL_ID=17721&URL_DO=DO_TOPIC&URL_SECTION=201.html> [Consulta: 06/06/2013].

35. OCDE, 2007.

36. Fernanda Peset, «Opinions del sector científic sobre la preservació d'informació» [en línia], *Blok de BiD*, 19/09/2012. <<http://www.ub.edu/blokdebid/ca/content/opinions-del-sector-cient%C3%ADfic-sobre-la-preservaci%C3%B3-de-la-informaci%C3%B3>> [Consulta: 06/06/2013].

2003/98/EC sobre reutilització de dades advocava per una major disponibilitat i incorporava les dades de recerca³⁷. A més, una comunicació oficial de l'any 2007 sobre l'accés, difusió i preservació de la informació científica a l'era digital equiparava les publicacions i les dades de recerca³⁸. Una comunicació posterior refermava la confiança en el potencial de les dades obertes per al creixement econòmic, l'avenç social, la transparència governamental i, naturalment, el progrés científic³⁹.

A la vegada, la UE ha donat suport actiu a les dades científiques obertes mitjançant els seus programes marc d'R+D, ja sigui a través de projectes d'investigació o de cofinançament d'infraestructures i dipòsits. En aquest darrer sentit, cal destacar les iniciatives OpenAIRE⁴⁰ i OpenAIREplus, que recullen entre les seves finalitats el tractament de dades científiques, l'enllaç amb les publicacions i la creació de serveis addicionals per als científics. Així, el mes de maig de 2013 ha entrat en funcionament Zenodo⁴¹ (Figura 3), un dipòsit digital obert impulsat per OpenAIRE i l'European Organization for Nuclear Research (CERN), amb el suport de la UE. Zenodo exemplifica una nova generació de dipòsits que impulsa la integració entre la producció científica, els conjunts de dades i la informació de finançament. Es tracta d'un servei senzill i innovador que permet compartir, distribuir, accedir i reutilitzar els resultats de recerca de tots els camps científics.

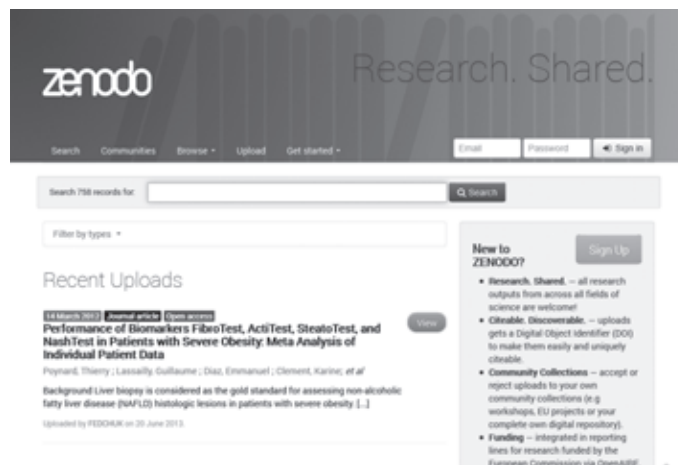


Figura 3. Zenodo

*Riding the wave: how Europe can gain from the rising tide of scientific data*⁴² és un informe europeu sobre els beneficis i els costos de la posada en funcionament d'una infraestructura global i multidisciplinària per treballar, compartir i conservar dades científiques. A l'informe, els experts es plantegen preguntes sobre com incentivar la contribució dels diferents agents (sobretot, cal implicar els investigadors), com garantir la preservació a llarg termini, com establir el context i origen de dades o, per últim, com sufragar la infraestructura.

37. UE, «Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information» [En línia], *Official Journal of the European Union*, 31/12/2003. <<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2003:345:0090:0096:EN:PDF>> [Consulta: 06/06/2013].

38. Comissió Europea (CE), *Communication from the Commission to the European Parliament, the Council and the European Economic and Social Committee on scientific information in the digital age* [En línia]: *Access, dissemination and preservation*, COM(2007) 56 final, Brussels: Commission of the European Communities, 2007. <http://ec.europa.eu/research/science-society/document_library/pdf_06/communication-022007_en.pdf> [Consulta: 06/06/2013].

39. Comissió Europea (CE), *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions* [En línia]: *Open data, an engine for innovation, growth and transparent governance*, COM(2011) 882 final, Brussels: European Commission, 2011. <<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2011:0882:FIN:EN:PDF>> [Consulta: 06/06/2013].

40. OpenAIRE: *Open Access Infrastructure for Research in Europe* (www.openaire.eu) [Consulta: 06/06/2013].

41. Zenodo (www.zenodo.org) [Consulta: 06/06/2013].

42. *Riding the wave: how Europe can gain from the rising tide of scientific data* [En línia], [S.l.]: European Union, 2010. <<http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf>> [Consulta: 06/06/2013].

L'any 2012 es van publicar els resultats d'una enquesta sobre el paper de la UE en l'accés i la preservació digital de la informació científica⁴³. Entre d'altres elements, l'enquesta posava de manifest algunes barreres per accedir a les dades científiques, com ara la falta d'inversió en la creació i manteniment d'infraestructures, pocs incentius als investigadors per fer disponibles les dades i manca de polítiques nacionals.

A l'Estat espanyol, l'article 37 de la Ley 14/2011, de 1 de junio, de la ciencia, la tecnología y la innovación, impulsa el desenvolupament de dipòsits digitals d'accés obert per a les publicacions. Així mateix, es fixa l'obligació de dipositar en accés obert la versió final dels continguts publicats que hagin estat finançats majoritàriament amb fons públics, però enlloc s'esmenten explícitament les dades científiques. Tot i això, sembla evident que les agències finançadores han de començar a implantar mandats obligatoris d'accés obert a les dades obtingudes amb diner públic, un cop els protocols i les infraestructures tecnològiques estiguin prou consolidades⁴⁴. En aquesta línia de treball, l'any 2012 la Fundación Española para la Ciencia y la Tecnología (FECYT) creà un grup de treball sobre «Depósito y Gestión de Datos en Acceso Abierto» en el marc del projecte RE-COLECTA, la primera iniciativa estatal per generar una infraestructura que impulsi la ciència oberta (*open science*).

Pel que fa a les editorials, les instruccions per als autors de revistes internacionals contempnen sovint la necessitat, i a vegades l'obligació, de compartir les dades de les recerques publicades⁴⁵. Una opció habitual és indicar a quin banc s'han de dipositar les dades, quins estàndards cal seguir o com s'han d'identificar unívocament els *datasets*. Si no existeixen dipòsits especialitzats, la mateixa revista pot facilitar l'emmagatzematge (per exemple, *Science* i, properament, *Nature* amb la plataforma *Scientific Data*). Si això tampoc no és possible, la revista té l'alternativa de reclamar



©istockphoto/Josef Kubicek

una disposició favorable per part dels autors a respondre peticions concretes. Ara bé, cal tenir present que l'adopció de polítiques editorials favorables al *data sharing* comporta algunes dificultats (per exemple, dades sensibles sobre pacients o fàrmacs), fet que pot repercutir en una rebaixa dels nivells d'exigència als autors. D'altra banda, els productors comercials de bases de dades comencen a incorporar grans conjunts de dades a la seva oferta de continguts. Així, Thomson Reuters va posar en funcionament l'octubre de 2012 el producte *Data Citation Index*⁴⁶.

Respecte a les institucions científiques, la seva responsabilitat principal radica a fixar una política interna de dades, amb la creació de dipòsits corporatius, l'establiment de guies de bones pràctiques i estàndards, la formació del personal o la coordinació dels diferents serveis implicats (informàtica, biblioteca i serveis científics)⁴⁷.

43. Comissió Europea (CE), *Online Survey on Scientific Information in the Digital Age* [En línia], Luxembourg: Publications Office of the European Union, 2012. <http://ec.europa.eu/research/science-society/document_library/pdf_06/survey-on-scientific-information-digital-age_en.pdf> [Consulta: 06/06/2013].

44. González Copeiro; Serrano-Muñoz, *op. cit.*, p. 42.

45. Torres-Salinas; Robinson-García; Cabezas-Clavijo, *op. cit.*, p. 180-181.

46. Data Citation Index (wokinfo.com/products_tools/multidisciplinary/dci) [Consulta: 06/06/2013].

47. González Copeiro; Serrano-Muñoz, *op. cit.*, p. 8.

Quina és la posició de la comunitat científica respecte a *data sharing*? Com a productors i usuaris de les dades, un 60% dels investigadors es declaren favorables a compartir-les, però la major part no ho porta a la pràctica: menys del 20% emmagatzema dades en un dipòsit digital o les envia als editors de revista conjuntament amb els manuscrits⁴⁸. Segons els investigadors, el principal obstacle per compartir és que no existeix una disponibilitat general de dades. A més, si estan disponibles, resulten difícils de localitzar o hi manca informació addicional per interpretar-les.

El projecte *Data Audit Framework* (DAF), finançat pel Joint Information Systems Committee (JISC) del Regne Unit, va pretendre esbrinar quin era el tractament de dades per part dels científics i com se'ls podia donar suport⁴⁹. La iniciativa detectà que els científics no disposaven d'una política clara per conservar les dades, les quals es trobaven disperses en diversos suports físics i formats informàtics, sense catàlegs de col·leccions de dades. Això comportava pèrdues irreparables, cerques per localitzar dades o identificar versions, etc.

Com a primer pas, el projecte DAF plantejava implementar polítiques corporatives per crear i conservar dades, amb el suport de guies per afavorir bones pràctiques: com tractar diferents tipus de dades, com fer còpies a llarg termini, com crear plans de gestió, etc. Altres propostes eren crear dipòsits institucionals i formar els nous investigadors en tècniques d'administració i preservació de dades. Ara bé, una de les principals dificultats derivava del complex cicle de vida d'un projecte de recerca, en el qual investigadors diferents podien recopilar dades de fonts diverses i en formats múltiples⁵⁰.

Segons els resultats del projecte DAF, en general els investigadors donaven poca importància a la reutilització posterior dels seus *datasets*. Malgrat això, cal tenir en compte que ells són els encarregats de proporcionar la informació bàsica per a la descripció de dades. Així, una descripció mínima ha de contemplar elements com el context de les dades (resum, objectius, metodologia de la investigació), descripció (contingut, estructura, terminologia, dates de creació i de modificació, versions, responsables), tecnologia (formats, estructura i nom dels fitxers, programes informàtics, sistemes d'emmagatzematge i seguretat) i, per últim, aspectes legals i polítiques d'accés⁵¹.

«Encara que hi ha una certa manca en els marcs tècnics i institucionals per regular la normalització de l'accés obert a les dades de recerca, de manera gradual estan emergint iniciatives internacionals que afavoreixen poder-les compartir»⁵². És necessària una anàlisi profunda sobre la identificació complexa dels drets d'autor d'objectes digitals que no són publicacions. Amb les possibilitats actuals de publicació en accés obert, els investigadors han d'establir drets i tipus de llicència perquè les dades puguin ser reutilitzades⁵³: Creative Commons, Scientific Commons, Open Data Commons, etc.

4. Documentació científica i gestió de dades

En el futur serà imprescindible comptar amb dipòsits de dades de recerca on dipositar i compartir els resultats de les investigacions pagades amb fons públics. «¿Quién se encargará de normalizar los datos, de describirlos, de man-

48. Borrego, 11/04/2012, *op. cit.*

49. Daniel Torres-Salinas, «Hacia la gestión de datos de investigación en las universidades: la *Data Asset Framework*», *Anuario ThinkEPI*, vol. 4 (2010), p. 262-265.

50. Àngel Borrego, «Els reptes de la gestió de dades de recerca» [en línia], *Blok de BiD*, 06/11/2012. <<http://www.ub.edu/blokdebid/ca/content/els-reptes-de-la-gesti%C3%B3-de-dades-de-recerca>> [Consulta: 06/06/2013].

51. González Copeiro; Serrano-Muñoz, *op. cit.*, p. 16.

52. García-García *et al.*, 2012, *op. cit.*

53. Arano *et al.*, *op. cit.*, p. 401.

tener y preservar esos repositorios de datos, de investigación, gubernamentales...? ¿También dejaremos que sean otros? ¿No es nuestra profesión la gestión de información, sea en bibliotecas, en centros de documentación o en empresas?»⁵⁴. En aquest context, les qualitats organitzatives i de preservació dels bibliotecaris-documentalistes són essencials per afrontar l'allau de dades científiques. La idea de treballar amb informació factual a les biblioteques no és una novetat, ja que des dels anys seixanta del segle XX hi havia *data librarians*, sobretot als Estats Units d'Amèrica⁵⁵. Aquest perfil presenta possibilitats excel·lents per establir relacions sòlides entre biblioteques i científics.

En un editorial de l'any 2009, la revista *Nature* ja reclamava l'equivalent digital de les biblioteques actuals per desenvolupar el *data sharing*, i apuntava directament cap a les biblioteques universitàries⁵⁶. D'altra banda, la Ligue des Bibliothèques Européennes de Recherche - Association of European Research Libraries (LIBER) impulsà un *Working Group on e-Science*. L'informe final del grup reuneix deu recomanacions per a les biblioteques per començar les tasques vinculades amb les dades científiques⁵⁷.

En general, la contribució de les biblioteques pot situar-se al voltant de tasques de descripció segons estàndards internacionals, polítiques de preservació digital i qüestions relacionades amb la reutilització i privacitat de dades⁵⁸. Per tant, s'obren bones perspectives de futur professional en aquest camp, on els nous rols i funcions dels gestors d'informació poden ser els següents:

- Des de les biblioteques: vetllar per la qualitat de la informació associada als conjunts de dades, implementar estàndards i metadades per a l'intercanvi d'informa-

En general, la contribució de les biblioteques pot situar-se al voltant de tasques de descripció segons estàndards internacionals, polítiques de preservació digital i qüestions relacionades amb la reutilització i privacitat de dades

ció i ontologies que ajudin a organitzar les dades per a la posterior recuperació, impulsar la interoperabilitat dels sistemes, garantir la preservació adequada dels dipòsits digitals.

- En relació als investigadors: localitzar conjunts de dades rellevants per a les investigacions, actuar com a intermediaris amb els dipòsits, formar els investigadors, redactar i difondre guies de bones pràctiques, promoure l'ús de llicències alternatives al copyright.
- Respecte als governs, entitats científiques i organismes internacionals: col·laborar en l'elaboració de polítiques, directrius o normes; donar suport a la implementació i manteniment d'infraestructures.

En aquest sentit, els professionals han de gestionar de forma integral el cicle de vida de les dades d'investigació

54. Tony Hernández-Pérez, «Datos abiertos y repositorios de datos: ¿nuevo reto para los bibliotecarios» [en línia], *Notas ThinkEPI*, 4 febrero 2013. <<http://www.thinkepi.net/datos-abiertos-repositorios-datos-nuevo-reto-bibliotecarios>> [Consulta: 06/06/2013].

55. Luis Martínez-Urbe; Stuart Macdonald, «Un nuevo cometido para los bibliotecarios académicos: data curation», *El Profesional de la Información*, vol. 17, núm. 3, p. 275.

56. Torres-Salinas, *op. cit.*, p. 261.

57. Birte Christensen-Dalsgaard (coord.), *Ten recommendations for libraries to get started with research data management* [en línia], The Hague: LIBER, 2012. <http://www.libereurope.eu/sites/default/files/WGSC_20120801.pdf> [Consulta: 06/06/2013].

58. Miquel Plaza-Navas; Agnès Ponsati Obiols; Isabel Bernal Martínez; Mercedes Baquero Arribas, «La biblioteca de recerca i la seva aportació al desenvolupament de l'activitat R+D+i», *Item*, núm. 56 (2012), p. 22.

amb la finalitat de garantir la seva preservació, localització i reutilització (*data curation*)⁵⁹ (Figura 4). El cycle ha de contemplar la recollida, selecció, integració, anàlisi, publicació i conservació de dades en un dipòsit. El desplegament acadèmic de la matèria, així com dels projectes de recerca relacionats, són factors determinants.

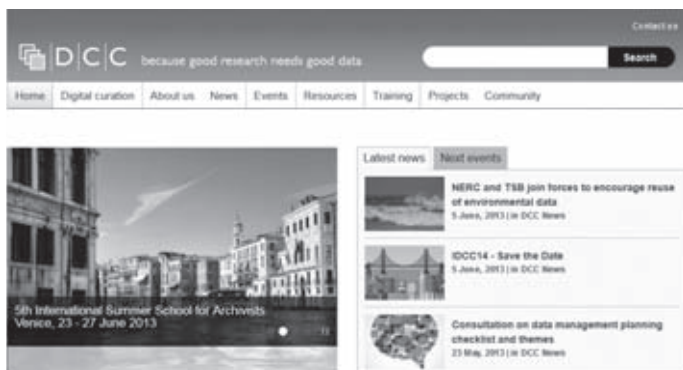


Figura 4. Digital Curation Centre (DCC)

La major part de les biblioteques considera la conservació de dades com a part del seu treball, tot i que encara són una minoria les que l'accepten per a l'emmagatzematge i la preservació⁶⁰. D'altra banda, el projecte europeu ODE ha estudiat la problemàtica associada a la citació de les dades, les quals permeten atorgar crèdits als autors dels conjunts, conservar evidències empíriques de recerques prèvies i facilitar als investigadors la localització i reutilització de dades⁶¹. Mitjançant una enquesta a bibliotecaris, el projecte afronta qüestions com, per exemple, quines metadades cal recopilar a la citació o com s'han de citar les actualitzacions.

A l'Estat espanyol es detecta una manca de sol·licitud de serveis bibliotecaris en relació a les dades científiques⁶². Per tant, s'imposa la formació adequada dels bibliotecaris, seguida de la promoció intensiva de les seves capacitats per generar confiança entre els investigadors.

5. Experiències en gestió de dades científiques

5.1. Directoris internacionals

En el camp emergent de la gestió i preservació de dades científiques, resulta evident la necessitat de comptar amb instruments d'identificació sistemàtica dels dipòsits temàtics i institucionals. Per tant, es requereixen directoris per identificar de manera fàcil i fiable els dipòsits especialitzats amb el propòsit de facilitar la descoberta de nous conjunts de dades, la recombinació dels quals, des d'una perspectiva multidisciplinària, fomentarà la innovació i la rendibilitat de la inversió en ciència⁶³.

Databib⁶⁴ (Figura 5) es defineix com un catàleg, registre, directori i bibliografia que s'orienta a identificar i localitzar dipòsits de dades de recerca. Originalment impulsat per l'Institute of Museum and Library Services (IMLS) dels Estats Units, el projecte es fonamenta en un sistema d'editors globals, en el qual no hi ha representants nacionals. La major part dels 575 dipòsits actuals corresponen a ciències biològiques (25,2%), ciències mediambientals (12,7%), ciències de la terra (10,6%), matemàtica i física (9,6%) i ciències mèdiques i de la salut (8,9%). Per àrees geogràfiques, es constata un domini aclaparador dels Es-

59. Digital Curation Centre (DCC) (www.dcc.ac.uk) [Consulta: 06/06/2013].

60. Borrego, 11/04/2012, *op. cit.*

61. Borrego, 06/11/2012, *op. cit.*

62. González Copeiro; Serrano-Muñoz, *op. cit.*, p. 42.

63. García-García *et al.*, 2012, *op. cit.*

64. Databib (databib.org) [Consulta: 06/06/2013].

tats Units (324 dipòsits) i del Regne Unit (67). La fitxa de cada dipòsit enumera, entre d'altres informacions, l'entitat responsable, país, data d'inici, resum, matèries i condicions d'accés, reutilització i dipòsit de dades.



Figura 5. Databib

re3data.org: Registry of Research Data Repositories⁶⁵ està finançat per la German Research Foundation (DFG). Es tracta d'un registre global amb la finalitat de promoure la cultura de compartir, donar accés i millorar la visibilitat dels dipòsits de dades científiques. La descripció de cada un dels 338 projectes actuals és exhaustiva, amb informació general, organitzacions responsables, polítiques, aspectes legals, estàndards, comentaris i versions.

ODISEA: International Registry on Research Data⁶⁶ és un projecte conjunt de cinc universitats espanyoles. Recull dipòsits internacionals que conserven conjunts de dades, materials addicionals d'articles i documents gràfics i multimèdia⁶⁷. La classificació dels 177 dipòsits actuals es basa en les àrees de coneixement de la base de dades Essential Science Indicators (ESI), de Thomson Reuters. Les disciplines amb més dipòsits són biologia molecular i genètica, biologia i química, farmacologia i toxicologia i,

Les disciplines amb més dipòsits són biologia molecular i genètica, biologia i química, farmacologia i toxicologia i, finalment, ciències de les plantes i els animals.

finalment, ciències de les plantes i els animals. L'anàlisi geogràfica també confirma el predomini clar dels països anglosaxons. El registre de cada dipòsit conté camps de directori i d'anàlisi: entitat responsable, àrea geogràfica, tipus i quantitat de dades, format i grau de compliment del protocol OAI-PMH.

5.2. Iniciatives a l'Estat espanyol

A l'Estat espanyol les infraestructures dedicades a les dades de recerca encara es troben en un estat inicial i els projectes en funcionament són escassos. A continuació es comenten quatre de les iniciatives més destacades.

Des de l'any 2004, el Centro Nacional de Datos Polares (CNDP)⁶⁸ (Figura 6) recull les dades en brut i les de calibratges instrumentals que els investigadors espanyols

65. re3data.org: Registry of Research Data Repositories (www.re3data.org) [Consulta: 06/06/2013].

66. ODISEA: International Registry on Research Data (odisea.ciepi.org) [Consulta: 06/06/2013].

67. García-García *et al.*, 2012.

68. Centro Nacional de Datos Polares (hielo.igme.es) [Consulta: 06/06/2013].

subvencionats amb projectes estatals han d'enviar-hi obligatòriament en el termini màxim de tres mesos. El CNDP disposa d'un protocol de remissió, emmagatzematge i difusió de dades, les quals s'acaben integrant en el Global Change Master Directory (GCMD)⁶⁹.



Figura 6. Centro Nacional de Datos Polares (CNDP)

Digital.CSIC⁷⁰ és el dipòsit digital institucional del Consejo Superior de Investigaciones Científicas (CSIC), que compta amb 72.642 registres (juny de 2013), dels quals 15 són conjunts de dades procedents de 9 instituts d'investigació de camps científics ben diferenciats: Estación Experimental de Aula Dei (EEAD), Instituto de Ciencias Marinas de Andalucía (ICMAN), Instituto de Estudios Documentales sobre Ciencia y Tecnología (IEDCYT), Instituto de Historia (IH), Instituto de Lenguas y Culturas del Mediterráneo y Oriente Próximo (ILC), Instituto de Química Física Rocasolano (IQFR), Instituto Mediterráneo de Estudios Avanzados (IMEDEA), Institut Milà i Fontanals (IMF) i, per últim, Instituto Pirenaico de Ecología (IPE).

En el camp de l'ensenyament superior, cal destacar la iniciativa de la Universitat Pompeu Fabra (UPF) d'integrar conjunts de dades a l'e-Repository⁷¹, el dipòsit digital de la institució⁷². Actualment presenta sis conjunts de dades procedents de l'Institut Universitari de Lingüística Aplicada (IULA) i tres del Departament de Ciències Polítiques i Socials.

La Biblioteca de Datos del Centro de Estudios Avanzados en Ciencias Sociales (CEACS)⁷³, que forma part de l'Institut Juan March de Estudios e Investigaciones, proporciona una col·lecció extensa de dades i d'informació agregada i geogràfica. També proporciona serveis especialitzats i formació als investigadors sobre temes d'opinió pública, sistemes polítics, eleccions, enquestes socials, etc.

6. Conclusions

En un context general d'obertura de quantitats massives de dades, considerades com actius amb valor afegit, és imprescindible una gestió adequada per garantir la seva disponibilitat i preservació a llarg termini. Amb la finalitat de generar nous coneixements, «una óptima gestión de [...] datos favorece la innovación y el desarrollo de la misma, puesto que permitiría la explotación de datos de alta calidad (compartir-reutilizar)»⁷⁴. Es tracta d'un domini emergent, amb bones perspectives de futur però també amb nombroses incògnites a resoldre. Per tant, els bibliotecaris-documentalistes han de col·laborar amb la resta d'agents implicats amb l'objectiu d'impulsar polítiques, directrius, infraestructures, estàndards, guies de bones pràctiques i accions de formació.

69. Global Change Master Directory (gcmd.nasa.gov) [Consulta: 06/06/2013].

70. Digital.CSIC (digital.csic.es) [Consulta: 06/06/2013].

71. e-Repository (repositori.upf.edu) [Consulta: 06/06/2013].

72. Arano et al., *op. cit.*, p. 387.

73. Biblioteca de Datos del Centro de Estudios Avanzados en Ciencias Sociales (www.march.es/ceacs/biblioteca/datalib) [Consulta: 06/06/2013].

74. González Copeiro; Serrano-Muñoz, *op. cit.*, p. 15.



©istockphoto/Sergey Skleznev

En el context del *big data*, l'àmbit de la investigació científica és un dels principals productors de dades. Les dades científiques tenen caràcter heterogeni i s'utilitzen com a matèria primera dels resultats, sobretot publicats en forma de documents. Cal mantenir el lligam de les dades amb les publicacions i possibilitar futures investigacions a partir de la reutilització i combinació de múltiples *data-sets*. En general, els investigadors es mostren favorables al *data sharing* però no el porten a la pràctica degut a la manca d'incentius, d'instruments tecnològics i de coneixements i habilitats sobre la matèria.

«Junto a una investigación de excelencia debe fomentarse también una gestión de los datos de excelencia, que permita su acceso óptimo y garantice su reutilización futura»⁷⁵. Per tant, en un futur immediat els bibliotecaris-documentalistes hauran de tenir un paper rellevant en la gestió integral del cicle de les dades científiques, des de la creació fins a la preservació. El desplegament acadèmic i professional del *data curation* facilitarà establir una relació

directa amb els investigadors i impulsar la promoció dels serveis bibliotecaris. En darrer terme, l'impacte del *data sharing* s'haurà de traduir en avenços científics i en més transparència i eficiència del sistema de recerca, però també en una major consolidació dels serveis d'informació i documentació científica.

Agraïments

A Mercè Martí i Robert Latorre, companys de l'Observatori de la Recerca (OR-IEC), per la revisió del text. A Netta Azgad, per la col·laboració en les traduccions.

75. Torres-Salinas, *op. cit.*, p. 265.

Bibliografía

ARANO, Silvia; MARTÍNEZ, Gemma; LOSADA, Marina; VILLEGAS, Marta; CASALDÀLIGA, Anna; BEL, Núria. «La comunidad “Recursos y datos primarios” de la Universitat Pompeu Fabra: los repositorios institucionales como infraestructuras científicas: estudio de caso». *Revista Española de Documentación Científica*, vol. 34, núm. 3 (2011), p. 385-407.

BERNERS-LEE, Tim. 5 *Star Open Data* [En línea]. <<http://5stardata.info>> [Consulta: 06/06/2013].

Biblioteca de Datos del Centro de Estudios Avanzados en Ciencias Sociales [En línea]. <<http://www.march.es/ceacs/biblioteca/datalib>> [Consulta: 06/06/2013].

BORGMAN, Christine L. «Data, disciplines, and scholarly publishing». *Learned Publishing*, vol. 21, núm. 1 (2008), p. 29-38.

BORREGO, Àngel. «Els reptes de la gestió de dades de recerca» [En línea]. *Blok de BiD*, 06/11/2012. <<http://www.ub.edu/blokdebid/ca/content/els-reptes-de-la-gesti%C3%B3-de-dades-de-recerca>> [Consulta: 06/06/2013].

BORREGO, Àngel. «Enriquir les publicacions amb dades empíriques» [En línea]. *Blok de BiD*, 11/04/2012. <<http://www.ub.edu/blokdebid/ca/content/enriquir-les-publicacions-amb-dades-emp%C3%ADriques-0>> [Consulta: 06/06/2013].

Centro Nacional de Datos Polares [En línea]. <<http://hielo.igme.es>> [Consulta: 06/06/2013].

CHRISTENSEN-DALSGAARD, Birte (coord.). *Ten recommendations for libraries to get started with research data management* [En línea]. The Hague: LIBER, 2012. <http://www.libereurope.eu/sites/default/files/WGSC_20120801.pdf> [Consulta: 06/06/2013].

COMISSIÓ EUROPEA (CE). *Communication from the Commission to the European Parliament, the Council and the Euro-*

pean Economic and Social Committee on scientific information in the digital age [En línea]: *Access, dissemination and preservation*. COM(2007) 56 final. Brussels: Commission of the European Communities, 2007. <http://ec.europa.eu/research/science-society/document_library/pdf_06/communication-022007_en.pdf> [Consulta: 06/06/2013].

COMISSIÓ EUROPEA (CE). *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions* [En línea]: *Open data, an engine for innovation, growth and transparent governance*. COM(2011) 882 final. Brussels: European Commission, 2011. <<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2011:0882:FIN:EN:PDF>> [Consulta: 06/06/2013].

COMISSIÓ EUROPEA (CE). *Online Survey on Scientific Information in the Digital Age* [En línea]. Luxembourg: Publications Office of the European Union, 2012. <http://ec.europa.eu/research/science-society/document_library/pdf_06/survey-on-scientific-information-digital-age_en.pdf> [Consulta: 06/06/2013].

Databib [En línea]. <<http://databib.org>> [Consulta: 06/06/2013].

Data Citation Index [En línea]. <http://wokinfo.com/products_tools/multidisciplinary/dci> [Consulta: 06/06/2013].

Digital.CSIC [En línea]. <<http://digital.csic.es>> [Consulta: 06/06/2013].

Digital Curation Centre (DCC) [En línea]. <<http://www.dcc.ac.uk>> [Consulta: 06/06/2013].

e-Repository [En línea]. <<http://repositori.upf.edu>> [Consulta: 06/06/2013].

FERRER-SAPENA, Antonia; SÁNCHEZ-PÉREZ, Enrique A. «Open data, big data: ¿hacia dónde nos dirigimos» [En línea]. *Notas ThinkEPI*, 1 febrero 2013. <<http://www.thinkepi.net/open-data-big-data-hacia-donde-nos-dirigimos>> [Consulta: 06/06/2013].

GARCÍA-GARCÍA, Alicia; RODRÍGUEZ-GAIRÍN, Josep-Manuel; SAORÍN, Tomás; GONZÁLEZ, Luis-Millán; GARCÍA-MASSÓ, Xavi; FERRER-SAPENA, Antonia; PESET, Fernanda. «ODISEA: International Registry on Research Data» [En línia]. *BiD: textos universitaris de biblioteconomia i documentació*, núm. 29 (2012). <<http://www.ub.edu/bid/29/garcia2.htm>> [Consulta: 06/06/2013].

GenBank [En línia]. <<http://www.ncbi.nlm.nih.gov/genbank>> [Consulta: 06/06/2013].

Global Change Master Directory [En línia]. <<http://gcmd.nasa.gov>> [Consulta: 06/06/2013].

GONZÁLEZ COPEIRO, Cristina; SERRANO-MUÑOZ, Jordi (coord.). *La conservación y reutilización de los datos científicos en España: informe del grupo de trabajo de buenas prácticas* [En línea]. Madrid: FECYT, 2012. <http://www.recolecta.net/buscador/documentos/informe_datos_cientificos_en_esp.pdf> [Consulta: 06/06/2013].

KEEFER, Alice. «La preservació de les dades de recerca i les agència de finançament de la R+D» [En línia]. *Blok de BiD*, 11/10/2011. <<http://www.ub.edu/blokdebid/ca/node/129>> [Consulta: 06/06/2013].

HERNÁNDEZ-PÉREZ, Tony. «Datos abiertos y repositorios de datos: ¿nuevo reto para los bibliotecarios» [En línia]. *Notas ThinkEPI*, 4 febrero 2013. <<http://www.thinkepi.net/datos-abiertos-repositorios-datos-nuevo-reto-bibliotecarios>> [Consulta: 06/06/2013].

MARTÍNEZ-URIBE, LUIS; MACDONALD, Stuart. «Un nuevo comedido para los bibliotecarios académicos: data curation». *El Profesional de la Información*, vol. 17, núm. 3, p. 273-280.

MELERO, Remedios. «Una plenamar de dades» [En línia]. *Blok de BiD*, 24/11/2010. <<http://www.ub.edu/blokdebid/ca/content/una-plenamar-de-dades>> [Consulta: 06/06/2013].

NATIONAL INSTITUTES OF HEALTH (NIH). *NIH Data Sharing Policy* [En línia]. Bethesda, Maryland: NIH, 2007. <http://grants.nih.gov/grants/policy/data_sharing> [Consulta: 06/06/2013].

ODISEA: International Registry on Research Data [En línia]. <<http://odisea.ciepi.org>> [Consulta: 06/06/2013].

OpenAIRE: Open Access Infrastructure for Research in Europe [En línia]. <<http://www.openaire.eu>> [Consulta: 06/06/2013].

ORGANITZACIÓ PER A LA COOPERACIÓ I EL DESENVOLUPAMENT ECONÒMIC (OCDE). *OECD Principles and Guidelines for Access to Research Data from Public Funding* [En línia]. Paris: OCDE, 2007. <<http://www.oecd.org/science/sci-tech/38500813.pdf>> [Consulta: 06/06/2013].

OPEN KNOWLEDGE FOUNDATION (OKFN). *The Open Data Handbook*. Cambridge: OKFN, 2012. <<http://opendata-handbook.org>> [Consulta: 06/06/2013].

PESET, Fernanda. «Opinions del sector científic sobre la preservació d'informació» [En línia]. *Blok de BiD*, 19/09/2012. <<http://www.ub.edu/blokdebid/ca/content/opinions-del-sector-cient%C3%ADfic-sobre-la-preservaci%C3%B3-de-la-informaci%C3%B3>> [Consulta: 06/06/2013].

PESET, Fernanda; FERRER-SAPENA, Antonia; SUBIRATS-COLL, Imma. «Open data y linked open data: su impacto en el área de bibliotecas y documentación». *El Profesional de la Información*, vol. 20, núm. 2 (2011), p. 165-173.

PIWOWAR, Heather A.; DAY, Roger S.; FRIDSMA, Douglas B. «Sharing Detailed Research Data Is Associated with Increased Citation Rate» [En línia]. *PLoS ONE*, vol. 2, núm. 3, e308. <<http://dx.doi.org/10.1371/journal.pone.0000308>> [Consulta: 06/06/2013].

Protein Data Bank (PDB) [En línia]. <<http://www.rcsb.org/pdb>> [Consulta: 06/06/2013].

RAMASAMY, A.; MONDRY, A.; HOLMES, C.C.; ALTMAN, D.G. «Key Issues in Conducting a Meta-Analysis of Gene Expression Microarray Datasets» [En línia]. *PLoS Medicine*, vol. 5, núm. 9, e184. <<http://dx.doi.org/10.1371/journal.pmed.0050184>> [Consulta: 06/06/2013].

re3data.org: Registry of Research Data Repositories [En línia]. <<http://www.re3data.org>> [Consulta: 06/06/2013].

PLAZA-NAVAS, Miquel; PONSATI OBIOLS, Agnès; BERNAL MARTÍNEZ, Isabel; BAQUERO ARRIBAS, Mercedes. «La biblioteca de recerca i la seva aportació al Desenvolupament de l'activitat R+D+i». *Item*, núm. 56 (2012), p. 13-32.

Riding the wave: how Europe can gain from the rising tide of scientific data [En línia]. [S.l.]: European Union, 2010. <<http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf>> [Consulta: 06/06/2013].

SWAN, Alma; BROWN, Sheridan. *To Share or not to Share: publication and Quality Assurance of Research Data Outputs* [En línia]. London: Research Information Network, 2008. <<http://www.rin.ac.uk/our-work/data-management-and-curation/share-or-not-share-research-data-outputs>> [Consulta: 06/06/2013].

TORRES-SALINAS, Daniel. «Compartir datos (*data sharing*) en ciencia: contexto de una oportunidad». *Anuario ThinkEPI*, vol. 4 (2010), p. 258-261.

TORRES-SALINAS, Daniel. «Hacia la gestión de datos de investigación en las universidades: la *Data Asset Framework*». *Anuario ThinkEPI*, vol. 4 (2010), p. 262-265.

TORRES-SALINAS, Daniel; ROBINSON-GARCÍA, Nicolás; CABEZAS-CLAVIJO, Álvaro. «Compartir los datos de investigación en ciencia: introducción al data sharing». *El Profesional de la Información*, vol. 21, núm. 2, p. 173-184.

UNESCO. *Charter on the Preservation of Digital Heritage* [En línia]. Paris: UNESCO, 2003. <http://portal.unesco.org/en/ev.php-URL_ID=17721&URL_DO=DO_TOPIC&URL_SECTION=201.html> [Consulta: 06/06/2013].

UNIÓ EUROPEA (UE). «Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information» [En línia]. *Official Journal of the European Union*. 31/12/2003. <<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2003:345:0090:0096:EN:PDF>> [Consulta: 06/06/2013].

Zenodo [En línia]. <<http://www.zenodo.org>> [Consulta: 06/06/2013].